PROJECT REPORT

NEW INDONESIA CAPITAL SENTIMENT ANALYSIS ON TWITTER USING SVM AND APPROACH WITH LEXICON METHOD

VARADA DWI CAHYA PUTRA
17.K1.0047

Faculty of Computer Science
Soegijapranata Catholic University
2022

APPROVAL AND RATIFICATION PAGE

Judul Tugas Akhir:	:	NEW INDONESIA CAPITAL SENTIMENT ANALYSIS ON TWITTER USING SVM AND APPROACH WITH LEXICON METHOD
Diajukan oleh	:	Varada Dwi Cahya Putra
NIM	:	17.K1.0047
Tanggal disetujui	:	15 Juli 2022
Telah setujui oleh		
Pembimbing	:	Y.b. Dwi Setianto S.T., M.Cs.
Penguji 1	:	Yonathan Purbo Santosa S.Kom., M.Sc
Penguji 2	:	Y.b. Dwi Setianto S.T., M.Cs.
Penguji 3	:	Hironimus Leong S.Kom., M.Kom.
Penguji 4	:	R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D
Penguji 5	:	Rosita Herawati S.T., M.I.T.
Penguji 6	:	Yulianto Tejo Putranto S.T., M.T.
Ketua Program Studi	:	Rosita Herawati S.T., M.I.T.
Dekan	:	Dr. Bernardinus Harnadi S.T., M.T.

Halaman ini merupakan halaman yang sah dan dapat diverifikasi melalui alamat di bawah ini.

sintak.unika.ac.id/skripsi/verifikasi/?id=17.K1.0047

DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : VARADA DWI CAHYA PUTRA

ID : 17.K10047

declare that this work, titled "NEW INDONESIA CAPITAL SENTIMENT ANALYSIS ON TWITTER USING SVM AND APPROACH WITH LEXICON METHOD", and the work presented in it is my own. I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at Soegijapranata Catholic University
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
3. Where I have consulted the published work of others, this is always clearly attributed.
4. Where I have quoted from the work of others, the source is always given.
5. Except for such quotations, this work is entirely my own work.
6. I have acknowledged all main sources of help.
7. Where the work is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Semarang, July, 15, 2022

VARADA DWI CAHYA PUTRA

17.K1.0047

APPROVAL PAGE FOR PUBLICATION OF SCIENTIFIC PAPERS FOR ACADEMIC INTEREST

The undersigned below :

Name : VARADA DWI CAHYA PUTRA

Undergraduate Program : TECHNICAL INFORMATION

Faculty : COMPUTER SCIENCE

Type of work : SKRIPSI

Approved to give Non-Exclusive Royalty Free Right to Soegijapranata Catholic University Semarang for scientific work entitled "New Indonesia Capital Sentiment Analysis On Twitter Using Svm And Approach With Lexicon Method" along with the existing tools (if needed). With this Non- Exclusive Royalty Free Right Soegijapranata Catholic University has the right store, transfer data / format, maintain and publis this final project as long as I keep my name as a writer / creator and as a Copyright owner.

This statement I made in truth

Semarang, July, 15, 2022

Sincarely

Varada Dwi Cahya Putra

17.K1.0032

ACKNOWLEDGMENT

Praise God Almighty who has given His mercy and grace to the author so that the author can complete the thesis entitled “New Capital Sentiment Analysis On Twitter Using Svm And Approach With Lexicon Method”. The preparation of this thesis aims to fulfill the requirements for obtaining a Bachelor of Computer Science degree at Soegijapranata Catholic University Semarang.

In preparing for this thesis, many parties have provided motivation, advice, and support to researchers. In this valuable opportunity, the researcher would like to express his gratitude and appreciation to all parties. First, I would like to thank my beloved parents for their endless love, prayers, and support. Thanks also to Mr. Y.B Dwi Setianto, S.T., M.Cs as supervising lecturer who has been willing to direct, provide input on ideas and the author’s guidance from the beginning to the end of this final project report.

The author would like to thank friends and good friends who have supported, provided input, ideas, motivation and enthusiasm for the author to complete this final project. Thanks also to Aldila Ghadah Nabilah, A.Md. Kes. who have accompanied and entertained the author in completing this final project.

ABSTRACT

Social media is growing fast on the internet. One of the most popular social media is Twitter. Many topics are discussed on Twitter such as economic, politic, social, culture, and law. One of the hot topics discussed on Twitter is the issue of relocating Indonesia's capital city. However, there is controversy from supporters and opponents. They have different views. This issue leads to a phenomenon of debate on Twitter that actually shows a collective concern about the public discourse. Sentiment analysis is a process of extracting, understanding and processing unstructured data to get sentiment information which is found in an opinion sentence. Application of sentiment analysis using machine learning methods shows that there are several method that are often used in

In this study, the Support Vector Machine (SVM) method is proposed to be applied to classified sentiment tweets on the topic of Indonesia new capital on social media twitter. The classification technique is carried out into 3 classes, namely positive and negative, and neutral. Before the classification process the data is labeled by lexicon method approach to help increase accuracy. This research also use K-Fold Cross Validation for Evaluation and Validation the classification model.

Based on testing on the sentiment of Indonesia new capital city from social media twitter from 3000 tweets (1674 positive, 1005, and 321 neutral) using SVM with lexicon method approach obtained accuracy 86%. After evaluation dan validation using K-Fold Cross Validation the accuracy increase from 87% to 88,23%.

Keyword: SVM, Sentiment Analysis, Lexicon Based, K-Fold Cross Validation, Twitter

TABLE OF CONTENTS

COVER ..... **i**

APPROVAL AND RATIFICATION PAGE ..... **ii**

DECLARATION OF AUTHORSHIP ..... **iii**

ACKNOWLEDGMENT ..... **v**

TABLE OF CONTENTS ..... **vii**

LIST OF FIGURE ..... **ix**

LIST OF TABLE ..... **x**

CHAPTER 1 INTRODUCTION **1**

1.1. Background 1

1.2. Problem Formulation 1

1.3. Scope 2

1.4. Objective 2

CHAPTER 2 LITERATURE STUDY **3**

CHAPTER 3 RESEARCH METOLODOGY **7**

3.1. Literature Study 7

3.2. Data 7

3.3. Coding 7

3.4. Analysis 7

3.4.1. Crawling Data 7

3.4.2. Pre Processing 7

3.4.3. SVM 8

3.4.4. K-Fold Cross Validation 8

CHAPTER 4 ANALYSIS AND DESIGN **9**

4.1. Analysis 9

4.1.1. Crawling data 9
4.1.2. Preprocessing 10
4.1.3. Determine Sentiment 12
4.1.4. TF-IDF 14
4.1.5. SVM Classification 14
4.2. Design 15
CHAPTER 5 IMPLEMENTATION AND RESULTS 16
5.1. Implementation 16
5.2. Results 20
CHAPTER 6 CONCLUSION 23
REFERENCES 24
APPENDIX ..... a

LIST OF FIGURE

Figure 4.1 Flow Chart 15

Figure 5.1 Crawling Result 20

Figure 5.2 Preprocessing Result 20

Figure 5.3 Sentiment Determination Result 21

Figure 5.4 Classification Report 21

Figure 5.5 K-Folding Cross Validation Result 22

LIST OF TABLE

Table 4.1. Example Data 9

Table 4.2. Preprocessing Example Data 11

Table 4.3. Sentiment Result..... 12

Table 4.4. Example Data of Sentiment Result..... 13